Open Research Online

The Open University's repository of research publications
and other research outputs

Spotlight browsing of resource archives

Conference or Workshop Item

oro.open.ac.uk

Spotlight Browsing of Resource Archives

Paul Mulholland, Trevor Collins and Zdenek Zdrahal
Knowledge Media Institute
The Open University
Milton Keynes, MK7 6AA, UK
+44 (0)1908 654506

{P.Mulholland|T.D.Collins|Z.Zdrahal}@open.ac.uk

ABSTRACT

Many organizations, particularly in the heritage sector, have large archives of digital content that they could make available to the general public or special interest groups if they had the appropriate mechanisms. Currently, these organizations can develop pre-crafted web sites, simple database-driven web sites or search facilities for accessing the content. However, none of these can be expected to appropriately present this content or scaffold its effective use.

Our proposed solution is an approach to navigation that we term *spotlight browsing*. It has the following key features: (i) Users can select a collection of resources from the archive, shining a spotlight on this area of the archive; (ii) The collection is structured in a number of ways to support its exploration and convey interesting properties of the collection; (iii) Users can see what is on the periphery of their current collection in order to encourage further exploration; (iv) Users can redefine the collection in order to move their spotlight to another area of the archive; (v) Any item viewed while browsing can be bookmarked into a personal collection that can be built up using resources from many different spotlights. The approach has been implemented and tested using an archive of content from a heritage institution.

Categories and Subject Descriptors

H.5.4 [Information Interfaces and Presentation]: Hypertext/Hypermedia - *architectures, navigation, user issues*

General Terms

Design, Human Factors

Keywords

Digital collections, browsing behavior, resource archives, content structuring, ontologies, narrative

1. INTRODUCTION

Many organizations have large digital archives of content that they would like to make available to the public or special interest

groups. This is true of many heritage custodians as well as others. These organizations could either pre-craft a web site of the available content, provide a search facility, develop simple database-driven access to the resources or some combination of these. However all of these have their drawbacks. A pre-crafted web site may involve an excessive amount of effort to construct and then only provide a particular perspective on the content. A simple database-driven site provides access to the content but little or no conceptual structure to guide its exploration. A search engine is good for accessing specific resources (and answering specific questions on the part of the user). Search is not so good if the user is trying to familiarize themselves with, and get an overview of, a collection of content. In this respect, our aim is similar to that of Chang et al [8], in helping the user to understand a collection of resources, rather than select a specific resource.

Our approach, which we term *spotlight browsing* has the following characteristics:

- Users specify a collection of interest from within the archive.

- The collection is internally structured to support its exploration.

- Some of the structure is created bottom-up from the properties of the content.

- Some of the structure is created top-down from the key messages that the content holder wishes convey.

- Top-down and bottom-up analysis supports the redefinition of the spotlight by pointing out what is on the periphery of the collection.

- A permanent collection allows the user to bookmark content from any particular spotlight they have created.

- The permanent collection can also be structured using similar mechanisms, pointing out patterns in the selected resources.

Unlike pre-crafted web sites or simple database-driven web sites, spotlight browsing provides multiple meaningful views of the content specified in the collection. Unlike search, spotlight browsing emphasizes learning and exploration across resources rather than the selection of specific items.

The rest of this paper is structured as follows. Section two outlines related work and some of the thinking behind the approach taken to spotlight browsing. Section three describes the context in which we implemented and tested spotlight browsing. Section four describes the technical approach we took. Sections five and six present a scenario of spotlight browsing in use and report on a preliminary evaluation. Discussion and future work are outlined in section seven.

2. RELATED WORK

Spotlight browsing is concerned with supporting the exploration of a collection of resources rather than searching and accessing specific resources. Our aim was to organize the collection in way that helps to highlight the meanings and relationships that exist across the collection as a whole. Our approach was influenced by the work of Pearce [23] that views a selected and organized collection of resources as a narrative that expresses a meaning beyond the individual resources making up the collection. The items within the collection constitute the vocabulary. Rules as to how they can be combined form the grammar. The rules and vocabulary together constitute the language with which the narrative can be expressed. For example, a curated exhibition of paintings constitutes a narrative that expresses a story across the selected works. A digital narrative could therefore be a collection of textual resources organized into a larger (hyper) text structure or pictures organized into a (virtual) art gallery.

Within our approach, we build collections out of units that are meaningful in their own right, i.e. they are lexia [18] and are meaningful in isolation. Examples include textual story passages, meaningful video clips and paintings.

The building and exploration of collections has long been considered an important part of creativity. Creativity can be understood as the modification (e.g. in terms of scope and/or structure) of a conceptual space in order to elicit a new perspective or approach to a problem or subject [4, 15]. Many theories and frameworks of creativity from early stage models emphasize the importance of building and organizing collections in order to inspire or provoke creativity. Many describe the creative process as going through a preparation phase concerned with gathering relevant information and resources [19]. More recently, Shneiderman's [28] framework for creativity comprises four activities, the first of these is concerned with collecting in order to learn from works stored in libraries, on the Web, and other places.

Modifications to existing conceptual structures, and the generation of creative ideas, are motivated by the selection, description and organization of existing resources. These techniques help to reveal important conceptual patterns or relations across the resources such as similarities, contrasts and causalities. For example, within art, selecting and reorganizing existing paintings is a common technique used to inspire new works [11]. Supporting the exploration of archives through the building and investigation of collections was therefore an approach we wished to pursue.

A number of existing systems provide support for the collection and organization of digital resources for teaching, learning and exploration. The Walden's Paths environment [12, 27] supports teachers in organising found web pages into a path for students to use. The path imposes a reading order and can also tell a story or express some relationship across the content. Software for building personal collections can provide additional support for the layout and presentation of resources. The Garnet digital library [7], following a spatial hypertext approach [26], allows users to spatially lay out found resources. A new collection of documents can then be "scattered" over the layout to generate clusters. TopicShop [3] and HunterGatherer [25] can both provide views of different granularity on collected resources.

Many search engine developers are also looking at how the results of a query can be automatically organised. Etzioni et al [14] describe the ongoing development of the KnowItAll search engine. The aims of the KnowItAll project are information extraction from the web and the presentation of query results in new ways to support the activity of the user. So far, their work has focussed on information extraction rather than the dynamic presentation of results in order to facilitate problem solving and exploration. Facilities toward this end, such as resource clustering are starting to appear in public search engines such as http://clusty.com. Building on this work we aimed to develop an approach that supported learning across as well as within resources, and provided a range of different organizational structures such as resource categories and pathways across resources.

A number of Adaptive Hypermedia (AH) systems have been developed that select and organise content according to user or learner interests. Adaptive Hypermedia employs an explicit user model to describe the knowledge, goals and interests of the user [5]. Adaptation can be made in terms of content selection, presentation and navigation. In educational domains AH has been used to guide individual learners through a curriculum structure based on their performance on online tests. AH can also be used to provide additional learner freedom within a hypertext space by adding additional horizontal links to an existing space [6]. Within our approach, rather than building a model of the user to guide content selection, organization and presentation, we aimed to organize the collection defined by the user in order that it can "talk back" to them, but did wish to explore the use of navigation techniques such as horizontal linking.

A number of tools have been developed to support the search and exploration of collections, particularly in the heritage domain. Hyvonen et al [16] use semantic web technology to support online museum exploration. Their system has two main features. First, semantic view-based search allows the visitor to select query terms from a number of views (i.e. taxonomies) of the content, such as a taxonomy of artefacts (e.g. paintings, sculptures) and a taxonomy of creators of the work. Second, a semantic recommendation system points visitors from any specific artefact to other similar artefacts (e.g. same creator, same type of artefact). We also wished to support the exploration of heritage resources, but focus more on collections of resources rather than individual resources as the key unit of exploration.

Rutledge et al [24] describe an approach to hypermedia generation in which material is organized into a presentation according to the semantic description of the components. Components may be pictures, or text fragments. The content is organized into semantically similar clusters which are then used to create a navigable presentation. Alani et al [1] describe the Artequakt project in which software was developed in order to construct narratives from lower-level sentence units. The sentences used in narrative construction are automatically extracted and annotated from the web. They are composed into narratives using domain specific templates, for example a bibliographic template contains sections for date and place of birth, major achievements, etc. Some coherence issues remain to be dealt with regarding the way the meaning of individual sentences can fit together to create a larger story. We also wished to compose navigable and coherent presentations out of existing resources but focus on the use of specified collections made of

meaningful lexia in order to allow us to build navigable structures without having to address the problem of low-level narrative coherence.

3. APPLICATION CONTEXT

Our experiment in the implementation and testing of a spotlight browser has been carried out in the context of a heritage center. We worked with Bletchley Park, a heritage center now part of Milton Keynes in the UK and concerned with the history of code breaking and computing. Previously Bletchley Park had been a wartime code breaking and intelligence center. During this time Alan Turing developed a mechanical decryption machine called the Bombe and Tommy Flowers developed Colossus, the valve-based semi-programmable computer. Bletchley Park became a heritage center in the early 1990's.

As is the case with many heritage centers, Bletchley Park has a large number of digital resources that they are unable to display during the physical visit. This includes over one thousand transcripts of interviews with people who worked at the Park and hundreds of historical accounts of what was happening at the Park month-by-month. Although this content cannot be properly exploited during the visit it could provide an interesting web site resource if its use is appropriately scaffolded.

The decision was made to develop a spotlight browsing service for this content to support a post-physical visit experience, where recent visitors could use the content to follow-up interests developed while at the Park. The system was designed so that visitors could express their interests by SMS text message and then explore content related to their message via the web site when they returned home. The reason for using text messaging was to get visitors to perform some activity while at Bletchley Park that would remind and encourage them to use to web-based service. The service is not-for-profit and does not generate any revenue for the heritage centre.
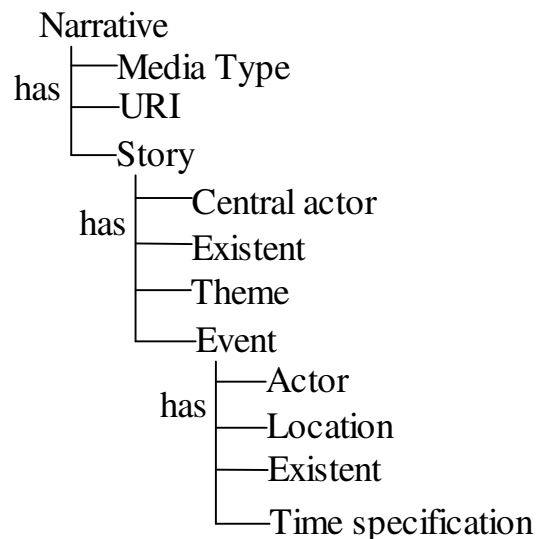
4. TECHNICAL APPROACH

The technical approach taken is built on our Story Fountain architecture for the semantic search, retrieval and organization of stories [22]. The selection and organization of the interviews and historical accounts is supported by their metadata description. For this we adopted the CIDOC Conceptual Reference Model (CRM) [10]. This is an ISO submitted standard ontology for the description of heritage domains. We used CIDOC CRM in order to facilitate the application of the software in other heritage domains.

We also developed an ontology of story and narrative for the description of the interviews and historical accounts and the concepts they contained. The story and narrative ontology followed structuralist theories of narrative in distinguishing between a story (the conceptualization of what is told) and a narrative (how it is told and what media is used) [9].

Each of the resources was annotated according to the CIDOC CRM ontology and story and narrative ontology. A narrative is the digital presentation of the story provided for the user. A narrative is described as having a media type (e.g. text, picture, audio), URI and associated story. A story is represented as having any number of central actors (e.g. main people or groups in the story), existents (e.g. main physical objects), themes and events. Each event itself is described as having actors, existents, locations and a time specification. Temporal information was described using Allen's time ontology [2]. Depending on the kind of event, existing properties were specialized or additional properties added. For example, an interview event had an interviewer and interviewee, and a creation event had a creator and object of creation. The overall metadata structure for describing an interview or historical account is shown in figure 1.

In order to support the selection of resources via a text message, a database thesaurus was developed. This maps textual terms (or near misses) to classes, properties, values or instances within the knowledge model. The terms entered in the text message are used to select a collection of resources to be structured for exploration. If a term maps to an instance (e.g. person, place, object) then all stories that have this instance as a value (either directly or in one of its associated events) are included in the collection. If the term maps to a class in the ontology (e.g. encryption machine) then all stories that have a value that is an instance of this class are selected. If the term is a property (e.g. creator) then all stories that have a value for that property are included in the collection.

Narrative
has —— Media Type
—— URI
—— Story
has —— Central actor
—— Existent
—— Theme
—— Event
has —— Actor
—— Location
—— Existent
—— Time specification

Figure 1. The overall structure for describing an interview or historical account.

This results in a collection of stories being retrieved according to the terms included in the text message. For example, the message "TURING BLOCKF" would cause the retrieval and organization of all stories that had either Alan Turing or Block F (one of the work areas in Bletchley Park when it operated as an intelligence and code breaking center) as a value of one of its direct slots (e.g. central actor or existent slot) or as a value of a slot of one of its associated events (e.g. actor or location slot).

Similar to the approach taken by Rutledge et al [24], the collection of stories retrieved for a set of terms can be clustered and categorized according to the overall set of properties and values in the collection of stories. This bottom-up organization of stories can introduce categories that are not directly related to the SMS terms but conceptually co-occur with them. This informs the user of additional concepts related to their query, that they may wish to find out more about. Organizing search results into

categories has also been found to increase the efficiency with which users can retrieve information from search results [13]. Also, Käki [17], in a longitudinal study of search engine use, found that categories are particularly useful when the user's search activities are exploratory or undirected as the categories scaffold the effective use of a larger number of search results.

As well as categories and clusters, the stories can also be organized into pathways according to the events that they contain. For example, two stories could be connected in a pathway from Alan Turing to Block F if one story contained an event that described how Alan Turing was associated with code breaking in Hut 8 (another work location) and a second story contained an event that described how work in Hut 8 was connected to Block F. This path-making facility was motivated by the Walden's Paths project [27] that has shown how paths connecting web pages can provide an educational and engaging way to explore content.

All of these organizations of the content are bottom-up, that is, they emerge from the metadata of the resources in the collection. By co-occurrence, categories are likely to be included that are not directly related to the terms in the message. For example, if the set of stories featuring Alan Turing are organized, the categories produced will feature concepts such as the Bombe and Hut 8. The large frequencies with which Bombe and Hut 8 feature in the collection of stories can be explained. Alan Turing was in charge of Hut 8 and developed the Bombe, therefore these concepts tend to co-occur. However, the bottom-up organization of the stories does not explicitly describe what the relationship is between these categories.

To solve this problem, the overall key message of Bletchley Park was formally represented as a concept network of nodes and links. The overall message was derived from the core presentation given by the tour guides. A portion of the concept network is represented in table 1.

Table 1. A portion of the concept network representation of the core Bletchley Park message.

Concept	Relation	Concept
Alan Turing	Was head of	Hut 8
Alan Turing	Developed	Bombe
Hut 8	Worked on naval messages with	Hut 4
Hut 8	Was headed by	Alan Turing
Bombe	Was developed by	Alan Turing

The concepts in the network (e.g. Alan Turing, Bombe) are also used in the metadata description of the stories. A set of relations (e.g. "was head of") were developed to express the tour guide story as a concept network.

Within the spotlight browser, the concept network is used to create horizontal links [6] between categories that are created bottom-up. For example, one link in the concept network expresses that Alan Turing developed the Bombe. Therefore, if a collection of stories features categories about both Alan Turing and the Bombe, this link will be shown whenever one of those story categories is viewed by the user.

Both bottom-up (from conceptual co-occurrence) and top-down (from the concept network) mechanisms are used to suggest what concepts could be added to the term set in order to support further exploration of the resource archive. For example, terms that co-occur with the current terms are suggested for inclusion in a future term set, as well as those that are explicitly linked to the current term set by the conceptual network of the tour.

Additionally, as well as iteratively defining different term sets and therefore moving the spotlight to different areas of the resource archive, users can bookmark any story to a permanent collection which can then be organized and explored using similar mechanisms.

Technically, the ontologies, metadata and concept network are represented in Lisp using the OCML modeling language [21]. Communication between Lisp and an Apache server is supported by Mod_Lisp [20]. Web page presentation is handled by scripts written in Python.

5. SCENARIO

A visitor to Bletchley Park would like to know more about Alan Turing and Block F. They send a text message stating their interests to the specified number. The terms in their message match concepts in the knowledge model and the collection of associated stories is retrieved.
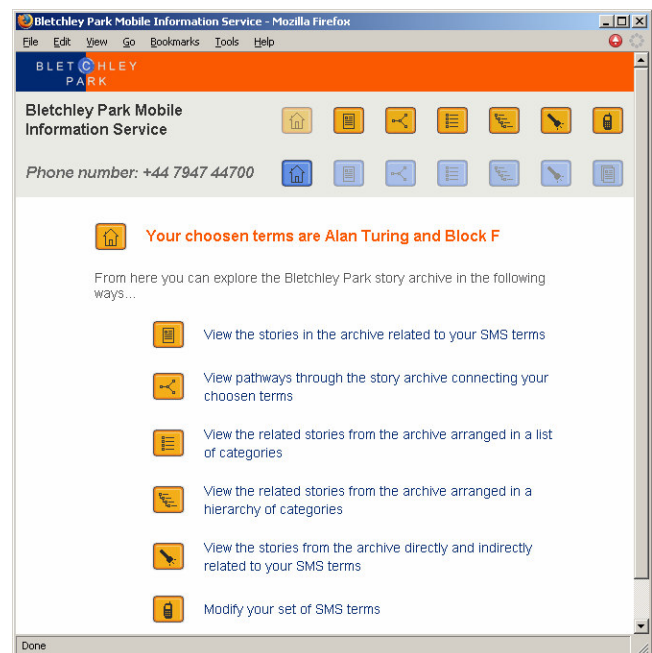

Figure 2. The options available for viewing the collection of stories. The final option allows the terms to be modified in order to create a new collection.

Later, from home, the visitor enters their phone number in the web site. They are then presented with the main page of the spotlight browser, which is entitled "Bletchley Park Mobile Information Service" (see figure 2). This page reminds them of the terms they submitted and provides a list of options for exploring the stories. The ordering of the options was chosen in order to

lead the user from viewing the stories according to their chosen search terms to viewing the stories according to other potential search terms not present in their query.

The first option (View the stories in the archive related to your SMS terms) provides a simple search facility for viewing stories that match specific SMS terms, for example, stories just about Alan Turing, stories just about Block F or stories that feature both.
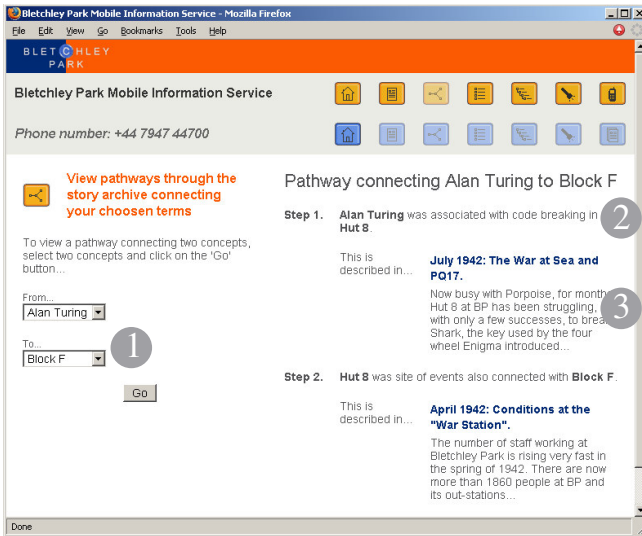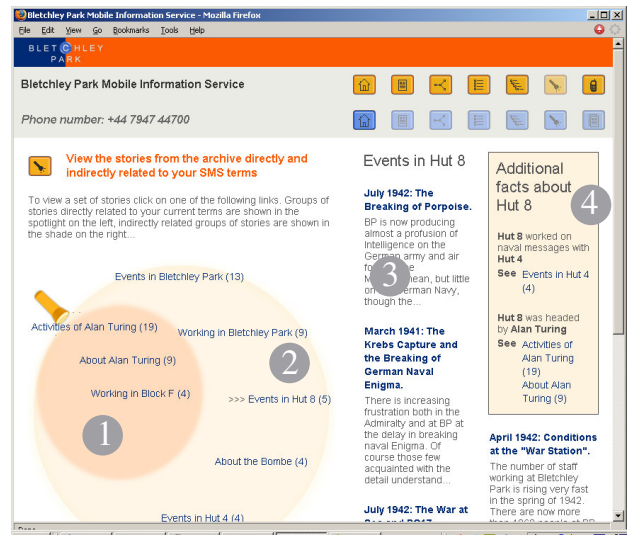

Figure 3. Making paths through the story collection. Key: (1) Selection of concepts to be connected; (2) A step in the path; (3) The story from which this step is taken.

Figure 4. Viewing the collection according to categories. Key: (1) The most common categories in the collection; (2) The set of stories in the selected category; (3) Horizontal links between this and other categories.

The second option available on the main page allows the visitor to view a linear path through the stories, connecting two of their search terms (see figure 3). The visitor finds out that Alan Turing and Block F are connected by two stories, one that links Alan Turing to Hut 8, and another that links Hut 8 to Block F. By clicking on the title of the story the visitor can find this out for themselves, as well as other issues covered in the stories.

The third option on the main page allows the visitor to view the story collection according to its most common categories (see figure 4). Many of these categories mention people, places and objects that were not covered by their original message, such as Hut 6, Bombe and John Tiltman. When viewing the stories from the category "Activities of Alan Turing", a shaded box provides additional facts about Alan Turing that make horizontal links to other available categories. One indicates that Alan Turing was the head of Hut 8, and the other that Alan Turing developed the Bombe. The categories and horizontal links provide additional information to the visitor and scaffold their reading of the stories.


Figure 5. Viewing categories in or on the periphery of the spotlight. Key: (1) Most common categories directly related to the current terms; (2) Categories not directly related to the current terms; (3) Stories in the selected category; (4) Horizontal links between this and other categories.

The fourth option allows the stories to be viewed hierarchically, similar to that of Rutledge et al [24], and also provides horizontal links between the categories. The fifth option organizes the stories in way that is intended to encourage the visitor to think about how they might modify their term set in order to continue exploration of the archive. Here the most common categories are organized as to whether or not they directly relate to the terms specified in the text message. This is illustrated by a graphical metaphor (see figure 5). Categories directly related to the text message are shown in the center of the spotlight created by the torch. Those not directly related to the message are shown on the periphery of the light. For example, "Events in Hut 8" and "About the Bombe" are significant categories that are not directly related to the terms in the text message. Once again the visitor can view the story

categories and the horizontal links between them as well as the stories themselves.
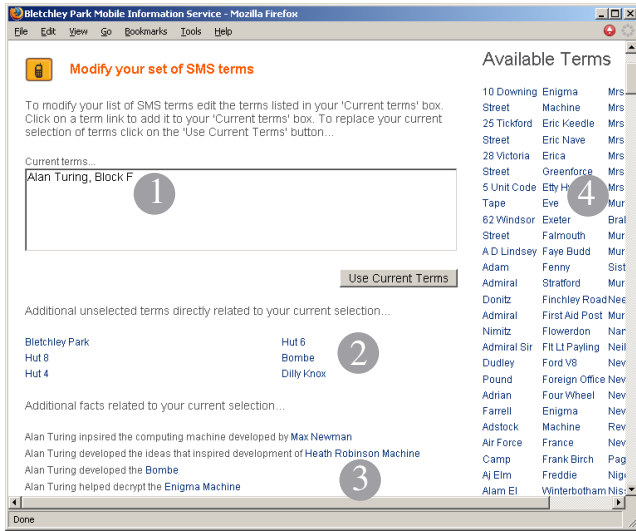
Figure 6. Modifying the chosen set of terms. Key: (1) The current term set; (2) Terms that most commonly co-occur with the current terms; (3) Terms explicitly linked to the current terms; (4) A few of the overall available terms.
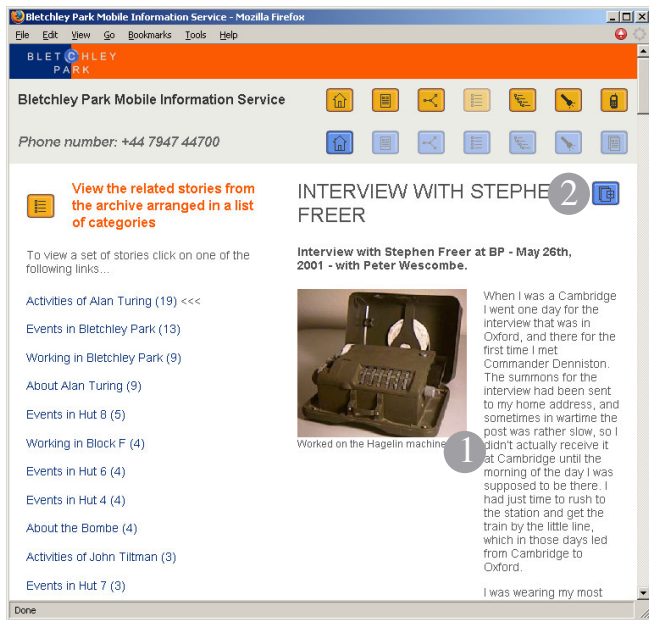

Figure 7. Viewing and book marking a story. Key: (1) The story itself; (2) The button to bookmark this into the permanent collection.

The sixth and final option allows the visitor to redefine their set of terms and start again with a new story collection (see figure 6). The visitor can manually delete anything from their current term box that they do not wish to explore further. Top-down and bottom-up

support is provided for the selection of new terms. Immediately below the term box, the six terms that most commonly co-occur with the current terms are provided. These can be clicked to add them to the current term box. Below them, further terms are provided that are connected to the current terms by a link in the concept network of the tour, such as "Alan Turing developed the Bombe". Clicking on any of these terms also adds them to the current term list. As may be expected, there is some overlap between the terms suggested bottom-up and top-down. Bombe appears in both lists of suggestions. Finally, the visitor can, if they wish, add further terms from the full term list, shown on the right hand side. Currently, there are over 800 such terms. Once a new term set has been chosen, the visitor can press the "Use Current Terms" button and return to the main page with a new story collection to explore.

A history of all the term sets used by the visitor is also recorded at the bottom of this page in case the visitor wants to recall a previous collection of stories and explore them once again.
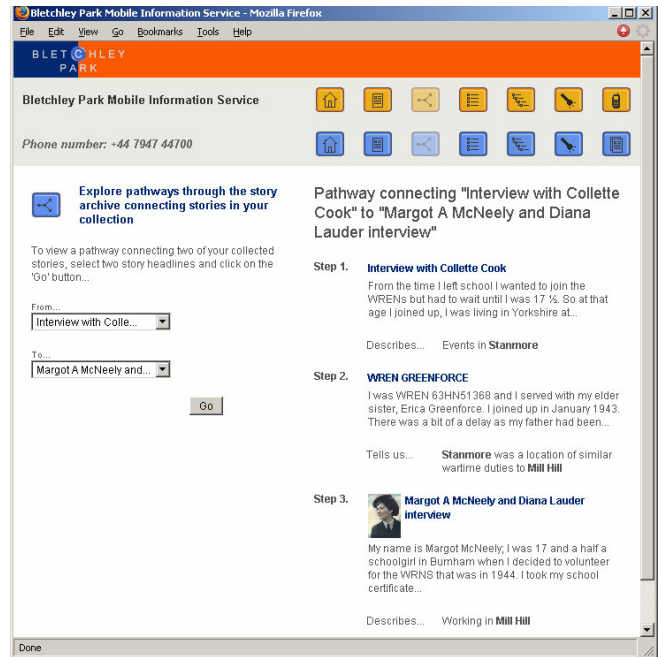
Figure 8. A pathway connecting two stories in a user's personal collection.

Whenever a visitor accesses a story from any of their collections generated by the set of terms, they can bookmark the story into a permanent collection. A story can be bookmarked by clicking on the blue icon in the right hand can corner of the story (see figure 7). The permanent collection can be explored by using the row of blue buttons rather than the row of orange buttons. Bookmarked stories can be explored using similar mechanisms (e.g. pathways, categories, hierarchies). For example, figure 8 shows a pathway between two stories added to the bookmarked collection. Note that this time the stories are selected to specify the start and end-points of the path rather than terms. This pathway consists of three stories. The first story describes events in Stanmore (a military center in an English town). The second story describes a connection between Stanmore and Mill Hill (another military

location near London). The third story describes working in Mill Hill.

6. PLEMINARY EVALUATION

A preliminary evaluation of the application was carried out with the aid of 35 students from a local high school and their teachers. The aim of the study was just to establish whether this kind of system could be used to explore an archive of content and identify usability problems.

At the beginning of the visit, all of the students were given a handout explaining the Bletchley Park Mobile Information Service and how to use it, including a list of about 150 terms that could be included in a message. The students spent about three hours exploring the Park either with a tour guide or studying the exhibits for themselves. During this time one of the teachers occasionally pointed out terms on the sheet related to an exhibit or what the guide had just said.

All the students sent one or more text messages related to people, places and objects mentioned by the guide or on the exhibits. At the end of the visit the students were reminded of how to access the site. Before leaving, six students volunteered to use the site from home, write a short essay on Bletchley Park and later provide feedback in an interview. For the rest of the students, accessing the site was purely optional and no classroom sessions were devoted to this activity. It was presented as an additional resource if students wished to follow up the subject in their own time.

During the following week, 20 of the students accessed the site, including the six who had previously volunteered. At the time of the study, the site did not provide support for bookmarking stories therefore only the orange buttons, and not the blue buttons were available on the web site. The essays provided by the six students demonstrated an ability to write a short essay coherently drawing on a number of different stories from the archive, including in some cases quotes from the stories. Other details were also included in the stories that had not been covered on the physical visit such as why certain codes were difficult to crack and background information on key characters. The topics of the essays included the ideas of Alan Turing, how the Bombe worked, and how naval messages were encrypted and decrypted.

A few weeks later the students were interviewed in a group by video-conference, by two of the authors. The students were asked a core set of questions. The interviewers then followed-up on their answers. The core questions were:

1. What interested you about Bletchley Park?
2. Why did you choose particular SMS terms?
3. Did you have any problems accessing the site?
4. Did you have any problems navigating the site?
5. What did you think of the stories?
6. What did you find out when using the web site?
7. Overall, if you were designing the site what would you do?
8. Did you use any other sources of information to write your essay?
9. Do you intend to find out any more about Bletchley Park?

In answer to the first question students pointed out a number of exhibits or stories told by the guide. For the second question the students explained that they texted people, objects and places that sounded interesting or they would like to know more about.

In answer to the third and fourth questions the students described a number of interface glitches and wording problems, that we had missed. Most, but not all, of these are have been fixed in the current version. In answer to questions five and six the students reported that they found some of the stories interesting and told us of a number of things they had found out about as a result of using the web site:

> "I knew that naval codes were difficult to crack but I didn't know there were different keys such as the shark key and the dolphin key."

> "I thought there was just one code and it was eventually cracked. I didn't realize they cracked a number of codes in turn."

> "I didn't know about Alan Turing's life before he came to Bletchley Park and how and were he lived when working at Bletchley Park."

> "I didn't know about the different dates and times when Colossus was used."

> "I didn't realize there was so much guess work involved in cracking the codes. They tried to guess the code that was being used and the words that the message might contain."

In response to question seven, the students offered interesting suggestions such as identifying and dealing with repetition across stories. For example, some stories, in terms of what they describe, are a subset of other stories. The stories that are a subset could be placed one level down in the navigation as additional reading. They also suggested the inclusion of more images which has yet to be addressed.

In answer to question eight, the students explained that they had also used other web sites or books about Bletchley Park. In response to question nine, one student asked for assurance that his phone number login would not expire as he wanted to continue to use the service. Two students said that they had since acquired books in order to follow-up further specific topics of interest. Another had persuaded her parents to take her back for a return visit. Although the web site cannot claim complete credit for motivating these future activities it clearly played a part in allowing them to follow-up in more detail topics of interest discovered during the physical visit.

Overall, the approach appeared to effectively scaffold the use of the resource archive and encourage learning and exploration across the resources. A more in-depth evaluation comparing our approach against standard browsing and search techniques is planned.

A number of interface features were redesigned following the evaluation. The software was then launched by Bletchley Park as a service for their visitors. To try out the current system, text "TURING BLOCKF" (quotes not required) to +44 (0)7985 400413. Then go to http://www.bletchleypark.org.uk/text and enter the number of the mobile phone you used to send the message (including the international code if not a UK number) and press GO.

7. DISCUSSION

We have presented an approach to scaffolding the use of a digital resource archive that we term *spotlight browsing*. Unlike search, the aim is to retrieve a collection of resources and explore across them rather than select specific resources for the focus of learning. Unlike standard browsing, additional structuring is dynamically provided for a particular collection of resources, including guidance on how the collection could later be redefined.

The key features of *spotlight browsing* are as follows:

- The user is able to define a collection for exploration. Definition can be by specifying concepts included in the stories (via the web site or text message) or explicitly by bookmarking individual items into a personal collection. By defining a collection the user is effectively shining a spotlight on a subspace of the available archive.

- The collection can be viewed from a number of different perspectives. Perspectives currently include pathways across the resources, common categories, hierarchies and categories inside and on the periphery of the user's spotlight.

- Top-down and bottom-up mechanisms are used to structure the content and scaffold its use. Bottom-up mechanisms highlight concepts that co-occur with those specified. Top-down mechanisms impose additional horizontal links derived from the core story that the content holder wishes to convey.

- Both iterative and permanent collections can be supported. The collection can be redefined many times to illuminate different parts of the archive. Resources of particular interest can be bookmarked into a permanent personal collection.

Although we have adopted a particular technical solution when implementing spotlight browsing, other approaches could be taken. For example, the selection of resources for a particular collection could be by example, where the user requests more stories that are similar to one they have found. Recommender systems could also be used to suggest additional content for a collection.

All of our approaches to the selection and organization of resources are driven by their semantic annotation. Other techniques, for example text analysis, could be used for the whole or partial selection and organization of content. Also, at the moment we have implemented a number of views on the collected content. Alternative structures could be used to organize the content, or alternative visual representations could be used to convey the structures that we currently use.

Our ongoing work and future work has five strands. First, we wish to look at dynamically altering or prioritizing the views provided depending on the nature of the collection. At the moment, the same views (path, category, etc.) are provided and instantiated regardless of the nature of the collection. However some views may be more or less effective depending on the size or other properties of the collection. Second, we wish to explore taking the user history into account when, for example, selecting the best views on the collection or suggesting how the collection could be defined. Our aim here would be to guide the user on a personalized trajectory across the archive, made up of the various spotlights on the content they make along the way.

Third, we would like to conduct a more detailed evaluation into the approach and its potential educational benefits. This would involve a detailed evaluation of the existing views on the collection we provide and ascertaining the possible benefits of each.

Fourth, we are yet to integrate the temporal annotations and reasoning mechanisms into the interface. This will allow users to filter or specify a collection according to a time span or temporally organize the contents of a collection. Fifth, we wish to test the approach with new kinds of content and scenarios of use. For example, we would be interested in using a different kind of content such as a photographic archive and other uses of the content, including more formal educational contexts.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] Alani, H., Kim, S., Millard, D., Weal, M., Hall, W., Lewis, P. and Shadbolt, N. Automatic ontology-based knowledge extraction from web documents. *IEEE Intelligent Systems, 18, 1* (2003), 14-21.

[2] Allen, J. Maintaining Knowledge about Temporal Intervals. *Communications of the ACM, 26, 11* (1983), 832-843.

[3] Amento, B., Terveen, L., Hill, W. and Hix, D. TopicShop: Enhanced Support for Evaluating and Organizing Collections of Web Sites. In *Proceedings of User Interface Software and Technology (UIST 2000)*, (San Diego, CA, 2000).

[4] Boden, M. Precis of The Creative Mind: Myths and Mechanisms. *Behavioural and Brain Sciences, 17, 3* (1994), 519-570.

[5] Brusilovsky, P. and Maybury, M. From adaptive hypermedia to the adaptive web. *Communications of the ACM, 45, 5* (2002), 30-33.

[6] Brusilovsky, P. and Rizzo, R. Map-Based Horizontal Navigation in Educational Hypertext. In *Proceedings of the ACM Conference on Hypertext and Hypermedia (HT '02)*, (College Park, Maryland, 2002).

[7] Buchanan, G., Blandford, A., Thimbleby, H. and Jones, M. Integrating information and seeking and structuring: Exploring the role of spatial hypertext in a digital library. In *Proceedings of the ACM Conference on Hypertext and Hypermedia (HT '04)*, (Santa Cruz, CA, 2004).

[8] Chang, M., Leggett, J., Furuta, R., Kerne, A., Williams, J. Burns, S. and Bias, R. Collection understanding. In *Proceedings of the Joint Conference on Digital Libraries (JCDL '04)*, (Tucson, Arizona, 2004).

[9] Chatman, S. *Story and Discourse: Narrative Structure in Fiction and Film.* Cornell University Press, New York, 1980.

[10] CIDOC Conceptual Reference Model. Available at http://zeus.ics.forth.gr/cidoc.

[11] Corrin, L. C., Kwon, M. and Bryson N. *Mark Dion.* Phaidon, London, 1997.

[12] Dave, P., Bogen, P. L., Karadkar, U. P. Francisco-Revilla, L., Furuta, R. and Shipman, F. Dynamically growing hypertext collections. In *Proceedings of the ACM Conference on Hypertext and Hypermedia (HT '04)*, (Santa Cruz, CA, 2004).

[13] Dumais, S., Cutrell, E. and Chen H. Optimizing search by showing results in context. In *Proceedings of the ACM SIGCHI conference on Human factors in computing systems (CHI '01)*, (Seattle, WA, 2001).

[14] Etzioni, O., Cafarella, M., Downey, D., Kok, S., Popescu, A., Shaked, T., Soderland, S. Weld, D. and Yates, A. Web-scale Information Extraction in KnowItAll. In *Proceedings of World Wide Web* (*WWW '04)*, (New York, 2004).

[15] Gero, J. S. Computers and creative design, In M. Tan and R. Teh (eds), *The Global Design Studio*, National University of Singapore, 1996, pp.11-19.

[16] Hyvönen, E., Junnila, M., Kettula, S., Mäkelä, E., Saarela, S, Salminen, M., Syreeni, A., Valo, A. and Viljanen, K. Finnish Museums on the Semantic Web: The user's Perspective on MuseumFinland. *Museums and the Web 2004*, (Arlington, VA, 2004).

[17] Käki, M. Findex: Search result categories help users when document ranking fails. In *Proceedings of the ACM SIGCHI conference on Human factors in computing systems (CHI '05)*, (Portland, Oregon, 2005).

[18] Landow, G. *Hypertext: The Convergence of Contemporary Critical Theory and Technology*. The Johns Hopkins University Press, Baltimore, 1992.

[19] Lubart, T. Models of the creative process: Past, present and future. *Creativity Research Journal, 13* (2000), 295-307.

[20] Mod_Lisp - Available at http://www.fractalconcept.com.

[21] Motta, E. *Reusable components for knowledge modelling: Case studies in parametric design problem solving*, IOS, Amsterdam, The Netherlands, 1999.

[22] Mulholland, P. Collins T. and Zdrahal, Z. Story Fountain: Intelligent support for Story Research And Exploration. In *Proceeding of the ACM Conference on Intelligent User Interfaces (IUI '04)*, (Madeira, Portugal, 2004).

[23] Pearce, S. M. *On Collecting: An Investigation into Collecting in the European Tradition*. Routledge, London, 1995.

[24] Rutledge, L., Alberink, M., Brussee, R., Pokraev, S., van Dieten, W. and Veenstra, M. Find the story - Broader applicability of semantics and discourse for hypermedia generation. In *Proceedings of the ACM Conference on Hypertext and Hypermedia (HT '03)*, (Nottingham, UK, 2003).

[25] schraefel, m. c., Zhu,Y., Modjeska, D., Wigdor, D. and Zhao, S. Hunter Gatherer: Interaction Support for the Creation and Management of Within-Web-Page Collections. *In Proceedings of World Wide Web (WWW '02)*, (Honolulu, Hawaii, 2002).

[26] Shipman, F. and Marshall, C. Spatial Hypertext: An Alternative to Navigational and Semantic Links, *ACM Computing Surveys Electronic Symposium*, 31, 4 (1999).

[27] Shipman, F., Furuta, R., Brenner, D., Chung, C. and Hsieh, H. Using paths in the classroom: experiences and adaptations. In *Proceedings of the ACM Conference on Hypertext and Hypermedia (HT '98)*, (Pittsburgh, Pennsylvania, 1998).

[28] Shneiderman, B. *Leonardo's Laptop: Human Needs and New Computing Technologies*. MIT Press, Cambridge, MA, 2002.